

“We Dissect Stupidity and Respond to It”: Response Videos and Networked Harassment on YouTube

American Behavioral Scientist
1–22

© 2021 SAGE Publications

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0002764221989781

journals.sagepub.com/home/abs



Rebecca Lewis¹, Alice E. Marwick²,
and William Clyde Partin²

Abstract

Over the past decade YouTube “response videos” in which a user offers counterarguments to a video uploaded by another user have become popular among political creators. While creators often frame response videos as debates, those targeted assert that they function as vehicles for harassment from the creator and their networked audience. Platform policies, which base moderation decisions on individual pieces of content rather than the relationship between videos and audience behavior, may therefore fail to address networked harassment. We analyze the relationship between amplification and harassment through qualitative content analysis of 15 response videos. We argue that response videos often provide a blueprint for harassment that shows both why the target is wrong and why harassment would be justified. Creators use argumentative tactics to portray themselves as defenders of enlightened public discourse and their targets as irrational and immoral. This positioning is misleading, given that creators interpellate the viewer as part of a networked audience with shared moral values that the target violates. Our analysis also finds that networked audiences act on that blueprint through the social affordances of YouTube, which we frame as harassment affordances. We argue that YouTube’s current policies are insufficient for addressing harassment that relies on amplification and networked audiences.

¹Stanford University, Stanford, CA, USA

²University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Corresponding Author:

Alice E. Marwick, Department of Communication, University of North Carolina at Chapel Hill, 115 Bingham Hall, Chapel Hill, NC 27514-3916, USA.

Email: amarwick@unc.edu

Keywords

harassment, YouTube, content moderation, networked audience, networked harassment, online video

Introduction

From 2017 to 2019, journalist Carlos Maza hosted Vox Media's *Strikethrough*, a left-leaning YouTube web series that critiqued mainstream media. *Strikethrough* was a frequent target of conservative YouTube pundits, particularly Steven Crowder, who hosts the popular, "politically incorrect" talk show *Louder with Crowder*. Crowder often posted critiques of *Strikethrough* that personally targeted Maza, who is Latinx and gay. In May 2019, Maza called attention to Crowder's harassing content when he tweeted a compilation of Crowder calling him a "lippy queer" and an "anchor baby," among other slurs. He wrote,

These videos get millions of views on YouTube. Every time one gets posted, I wake up to a wall of homophobic/racist abuse on Instagram and Twitter. . . . These videos makes [sic] me a target of ridiculous harassment, and it makes life sort of miserable. I waste a lot of time blocking abusive Crowder fanboys, and this sh** derails your mental health.

Maza claimed that YouTube had ignored his repeated reporting of Crowder's channel for harassment. His thread was viewed over 75,000 times and sparked a public discussion about the problems of harassment, free speech, and monetized political content on the platform (Alexander, 2019a).

After Maza's tweet went viral, YouTube responded with a series of inconsistent policy decisions, first declining to act against Crowder, then removing his ability to garner advertising revenue, then offering Crowder the ability to remonetize if he removed links to homophobic merchandise on his page. The inconsistency of YouTube's response revealed limitations in both the company's current policies and its ability to enforce them. YouTube determines violations of community guidelines by examining individual videos, and the specifics of content violations are often unclear. YouTube's community guidelines state that "we remove content promoting violence or hatred against individuals or groups" based on race or sexual identity; however, in the wake of Maza's Twitter thread, YouTube clarified that "opinions can be deeply offensive, but if they don't violate our policies, they'll remain on our site" (TeamYouTube, 2019).

Setting aside the spurious nature of this distinction, we argue that moderation policies based on individual pieces of content may miss the forest for the trees: the amplifier effect of a popular social media account shaming an individual. Networked harassment describes a form of online harassment against a target or set of targets which is encouraged, promoted, or instigated by members of a network, such as an audience or online community (Marwick & Caplan, 2018). This type of harassment may include name-calling, insults, threats, nonconsensual image sharing, and publicizing private information, but is distinguished from other forms of harassment because it happens at *scale*. As a result, experiences of online harassment, including

networked harassment (Madden et al., 2018; Sobieraj, 2018), rarely map to legal and technical models of harassment, which typically presume a dyadic model in which one person repeatedly engages in unwanted contact with another (Citron, 2014; Pater et al., 2016). In contrast, there may be hundreds or even thousands of participants in networked harassment campaigns, each of whom may only contact the victim once or twice. Moreover, high-profile accounts usually avoid explicitly directing their audience to harass a target, given that this is against most platforms' terms of service. This makes identifying and regulating networked harassment difficult using dyadic harassment models. This article therefore asks: What is the relationship between amplifiers and harassment?

To analyze this relationship, we examine the genre of "response videos." Response videos are made by YouTubers of one political valence as rebuttals to videos espousing an opposing political valence. Creators selectively edit and respond to the original video's claims, critiquing their arguments and levying personal attacks against their hosts. This remix of existing and original content enrolls the target into a kind of simulated debate, reframing what might have otherwise been a monologue into a back-and-forth exchange. This content structure interpellates the target as if they were already defending themselves against response video's creator and gives the respondent the last word.

Even without explicit calls for harassment, we argue that response videos create conditions in which harassment flourishes—a blueprint for harassment. If perpetrators of harassment often believe that their actions are justified (Blackwell et al., 2018; Jhaver et al., 2018), and if harassment is understood as punishment for norm violation (Crockett, 2017; Hill & Johnson, 2020), then these videos draw on social norms shared by a networked audience and the technical affordances of YouTube to explain to the audience not only why the target is wrong, but why they are immoral and why harassing them would be justified.

Literature Review

Networked Harassment

"Online harassment" is an umbrella term that encompasses behaviors ranging from name-calling to cyberstalking. Early studies of online harassment focused on teenagers and young adults bullying others, including spreading rumors, posting cruel comments, or making threats (Finkelhor et al., 2000; Tokunaga, 2010). Such studies usually involved participants younger than 18 years and harassment took place in front of a networked audience of peers on social media (Marwick & boyd, 2014). Recent work has expanded beyond this context. In a survey of more than 3,000 Americans, Lenhart et al. (2016) defined online harassment as a constellation of 10 behaviors, including name-calling, physical threats, sexual harassment, intimate and partner violence, and instigating harassment. Online harassment also includes publicizing personal information, nonconsensual image sharing, and contacting relatives, friends, or employers to harm the target (Sobieraj, 2018).

Feminist scholars situate online harassment within a broader landscape of misogyny and sexual violence (Henry & Powell, 2015; Jane, 2014). While diverse populations experience harassment, women, especially women of color and queer women, are far more likely to experience serious online harassment like stalking (Lenhart et al., 2016; Vitak et al., 2017), which discourages marginalized populations from participating online (Hill & Johnson, 2020; Moloney & Love, 2018). Sarah Banet-Weiser and Kate Miltner (2016) refer to “networked misogyny” as “an especially virulent strain of violence and hostility towards women in online environments.”

Marwick and Caplan (2018) define *networked* harassment as collaborative attacks against individuals amplified through social media. In other words, networked harassment involves the same behaviors defined by Lenhart et al (2016) and Sobieraj (2018) but at a much larger scale, involving hundreds or even thousands of participants within a short period of time. Importantly, these practices are enabled and shaped by the affordances of digital media, particularly social media platforms. On these platforms, networks of individuals bound by a common digital infrastructure can take part in spontaneous yet collaborative activities, with or without formal organization or planning. During GamerGate, for example, 4Chan users set up a dedicated IRC channel to organize harassment of game developer Zoe Quinn and activist Anita Sarkeesian, which they then carried out on Twitter (Chess & Shaw, 2015). In this case, a loose-knit network of individuals met on 4Chan (an anonymous platform hospitable to trolling and antifeminist views), moved to a chat server that afforded more private, real-time conversation, and used Twitter to recruit more participants and attract attention to the campaign. Unlike harassment that stems from a single individual’s actions, networked harassment is often spurred by the affective and social ties between individuals, what Zizi Papacharissi (2016) calls “mediated feelings of connectedness,” as participants can feel that they are taking part in a broader communal activity.

Given that perpetrators of harassment often believe their actions are justified (Blackwell et al., 2018; Jhaver et al., 2018), Marwick and Caplan (2018) argue that networks like the Men’s Rights Movement portray feminists as antimale to justify attacks against them. This promotes moral outrage, which people experience when they believe a moral norm has been violated, motivating them to shame and punish the violators (Crockett, 2017). Moral outrage is more common online than off, and often spreads through like-minded networks (Brady et al., 2017).

Based on these common elements, we conceptualize *networked harassment* as a technique of moral enforcement that spreads through ideologically bounded groups and is typically amplified by high-profile accounts, individuals, or organizations, which may be leveraged by actors across the ideological spectrum. In *networked harassment*, unlike cyberstalking or other dyadic models of harassment, an individual node is harassed by a group of people informally organized through social media, colloquially referred to as “dogpiling,” “brigading,” or “raiding.” While harassment can take place through a variety of technologies and platforms, we focus on response videos posted to the video-sharing platform YouTube. Although most research on harassment focuses on right-leaning groups, conceptualizing networked harassment as moral

enforcement allows left-leaning practices such as “callouts” and “putting people on blast” to become legible as networked harassment.

YouTube’s Affordances and Platform Accountability

YouTube’s affordances,¹ or the material qualities of technologies that shape (though not necessarily determine) the behaviors of their users, have influenced the practices and cultures on the platform. Because anyone with a camera phone and internet access can upload content to the site, YouTube has become a key platform for prospective influencers and online celebrities to develop and monetize their audiences (Abidin, 2018; Duffy, 2017). Such users often adopt strategies of “micro-celebrity,” in which creators carefully assemble a form of self-presentation “to be consumed by others,” and treat their audience as a fan base (Marwick, 2015, p. 337). Importantly, micro-celebrity practice involves direct interaction with this audience, creating a sense of shared intimacy (Berryman & Kavka, 2017; Raun, 2018).

YouTube also provides social functionality to both audiences and creators, facilitating the development of active, networked groups (Abidin, 2015; Cunningham & Craig, 2017; Lange, 2007, 2009). Because all YouTube users can view, comment on, share, and create content, distinguishing between creators and audiences is difficult. These affordances facilitate a “YouTube community” of native platform creators and viewers who interact with each other and each other’s content (Burgess & Green, 2018). This mix of hosting, broadcasting, and social affordances makes YouTube a productive space for “remix culture,” in which content creation is generative and adaptive (Fagerjord, 2010). Overall, the YouTube community involves a constant negotiation and set of discussions flowing from creators to audiences, audiences to creators, and between multiple creators and their respective audiences.

The implications of YouTube’s participatory culture are frequently treated optimistically by the company, the media, and, at times, academia. Concurrently, scholarship has examined the potential of participatory culture for enabling harassment, propaganda, and disinformation. As early as 2007, Patricia Lange investigated the broad phenomenon of “antagonism on YouTube,” challenging links between internet antagonism and anonymity by showing that various “hating” behaviors came from highly visible creators and audiences. Indeed, Burgess and Green (2018) argue that conflict, trolling, and harassment have been integral to YouTube since its origins. They write that prominent YouTubers “expressed reluctance to moderate or ban comments because those kinds of controls are counter to the ethos of openness that supposedly distinguishes participatory culture” (p. 119). Moreover, flame wars and face-offs between YouTubers and, at times, their fanbases are a lucrative form of entertainment on the platform (Burgess & Green, 2018, p. 119). While this antagonism is often political in nature, the boundaries between right/left ideologies are not always clear. For example, Stephen Pihlaja (2014) explored late 2000s “flame wars” between atheist and Christian YouTubers, in which videos acted as pieces of dialogue between warring groups. While the atheists considered themselves more progressive than Christian YouTubers, several later shifted their channels’ focus to criticizing feminism.

Sometimes this antagonism is outside the realm of politics altogether, such as feuds between YouTubers specializing in beauty products.

Social platforms like YouTube have come under criticism for their role as intermediaries in hosting disinformation, hate speech, and harmful content. Groups as diverse as breastfeeding activists, conservative pundits, and Black Lives Matter protestors have accused platforms of violating free speech by moderating content (Gillespie, 2018). Since social platforms are private companies, they are not governed by the First Amendment. Section 230 of the Communication Decency Act provides platforms a “safe harbor,” which makes them immune to prosecution for user actions. However, given the strong normative value of free speech, technology companies frequently uphold free speech as a primary value, while social actors leverage free speech discourse to argue against their content’s removal (Klonick, 2017). In 2019, responding to the systematic “de-platforming” of right-wing provocateurs like Alex Jones and Milo Yiannopoulos, Senate Republicans led a subcommittee hearing on “Stifling Free Speech: Technological Censorship and the Public Discourse” which interrogated Facebook, Twitter and Google on their presumed liberal bias against conservative media (Newton, 2019). However, two studies found no difference in engagement on Facebook between right-wing and left-wing political content (Martinez, 2019), while network analysis shows that right-wing content has flourished online (Benkler et al., 2018). Regardless, content moderation has become a partisan political issue and platforms have struggled to remove content that violates their Terms of Service while dodging accusations of censorship.

Research Site and Method

Research Site

The response video is a staple YouTube format in which one creator speaks about another creator’s content directly to their audience, sometimes inviting them to react as well. The format, which exemplifies remix culture, originates from a feature which enabled users to directly post a “video response” to another video (Burgess & Green, 2018). YouTube eventually phased out this feature, citing low click-through rates, but the practice remains an integral part of YouTube’s culture (The YouTube Team, 2013). Response videos are employed across various subcultural contexts, but we focus on political YouTubers.

We further focus our analysis on “anti-SJW” content (“social justice warrior,” right-wing slang for “liberal” or “progressive”), a genre that emerged in the early 2010s. Anti-SJW videos have thrived as part of a larger ecosystem of conservative, libertarian, and far-right media on YouTube (Lewis, 2018), with strong links between the so-called “alt-lite,” “Intellectual Dark Web,” and more extreme “alt-right” content (Ribeiro et al., 2020). While we do not argue that harassment is a primarily right-wing phenomenon, we focus on anti-SJW response videos for two reasons.

First, quantitative studies demonstrate that the landscape of online information is ideologically asymmetrical. Benkler et al. (2018) found that online polarization in

news sharing was heavily skewed to right-wing and far-right sources, and YouTube-specific studies demonstrate that right-wing videos are more popular than their left-wing counterparts (Soldner et al., 2019). Moreover, a comparative study of right-wing and “baseline” YouTube videos found that right-wing videos used more hateful vocabulary, particularly to describe Muslims, immigrants, and LGBTQ (lesbian, gay, bisexual, transgender, and queer) people (Otoni et al., 2018). Finally, far-right online spaces like 4chan heavily target YouTube for networked harassment, resulting in increased hate speech in YouTube comments (Hine et al., 2017).

Second, the anti-SJW YouTube community emerged largely as part of the Gamergate controversy in the summer of 2014. The incident involved gaming enthusiasts and antifeminist activists using “the strategies and tactics of participatory culture to target feminist media critics and game developers” (Marwick & Lewis, 2017, p. 8). While Gamergate took place across platforms, YouTube was a crucial hub for content creation (Burgess & Matamoros-Fernández, 2016) and many creators trace their rhetoric around issues of feminism and social justice to this incident. Thus, not only is YouTube a key site for far-right discourse, anti-SJW YouTubers are linked both historically and in the current moment to networked harassment.

Method

We performed a qualitative content analysis of a representative sample of 15 response videos uploaded between 2014 and 2018 by 15 different YouTubers, 13 right-wing and 2 left-wing (see Table 1). This sample was developed by viewing a range of response videos across political channels on YouTube. Following previous work on harassment, we focused on videos produced by the Alternative Influence Network, an informal, collaborative network of antiprogressive YouTubers (Lewis, 2018). Our sample was chosen to represent different types of Alternative Influence Network creators with different targets and approaches. We also included two left-wing videos in our sample as a point of comparison, and because a newer generation of left-wing creators has emerged in response to antiprogressive YouTubers (Amin, 2019). While our sample reflects the ideological bias of response videos overall, we recognize the need for further research on left-leaning creators and networked harassment.

Each video in our analysis was viewed, transcribed, checked for accuracy, and uploaded to qualitative data analysis software. The authors created a draft codebook based on watching the entire corpus and pilot coding three transcripts, and then coded all 15 videos using a grounded theory approach (Strauss & Corbin, 1990), focusing on thematic and stylistic elements and iterating on the codebook over time. We supplemented this coding by viewing extratextual content such as videos and tweets made by targets in response to harassment they received. Finally, we examined relevant video metadata, including “About” sections, likes/dislikes, views, and comments.

Findings

While our corpus contains no instances of creators explicitly advocating harassment among their audience, we argue that response videos nonetheless provide a *blueprint*

Table 1. Videos in the Corpus.

| Title | Creator | Target | Length (min.) | URL | Views (March 2020) |
|--|------------------------|------------------------------------|----------------|--|------------------------|
| Sensitive Tranny Thinks Words Are Violent | Blaire White | Riley Dennis | 6:49 | https://www.youtube.com/watch?v=31i_YpB7HIQv4 (removed by YouTube) | 629,000 |
| Racist Costumes With Franny | Andy Warski | Francesca Ramsey | 10:00 | https://www.youtube.com/watch?v=UpJk4e9m2k4 (removed by YouTube) | 145,000 |
| Dear Rebecca Watson who wants to keep talking about herself | Sargon of Akkad | Rebecca Watson | 12:19 | https://www.youtube.com/watch?v=fvmZleh_e3M | 254,000 |
| Feminism vs. FACTS (Anita Sarkeesian DESTROYED!) | Thunderf00t | Anita Sarkeesian | 37:28 | https://www.youtube.com/watch?v=i9JjU-1I1DTU | 1 Million |
| VOX REBUTTAL: Comedians vs. Trump? The Real Absurdity . . . | Steven Crowder | Carlos Maza | 11:15 | https://www.youtube.com/watch?v=ZmQs0K-RKQ (removed by YouTube) | 1.3 Million |
| Trans YouTuber Says It's UNFAIR He Can't Have BABIES | James Allsup | Gigi Gorgeous | 13:49 | https://www.youtube.com/watch?v=kuzFaub4UM (removed by YouTube) | 804,000 |
| Queer Kid Stuff: IT'S TIME TO STOP! | Hunter Avalone | Lindsay (Queer Kids Stuff) | 12:29 | https://www.youtube.com/watch?v=Y3k09OT3IBY (removed by YouTube) | 298,000 |
| It's Okay To Celebrate Thanksgiving FEMINISM IS FOR POOPYHEADS!!! Re: Kat Blaque | No Bullshit Shoe0nhead | Imaeyan Ibanga (A++) Kat Blaque | 17:25 13:04 | https://www.youtube.com/watch?v=7Y-R6qFjTYM https://www.youtube.com/watch?v=iqVNElpCEko | 174,000 1.5 million |
| Gay Muslim Drag Queen Causes Confusion | Paul Joseph Watson | glamrou/BBC | 6:56 | https://www.youtube.com/watch?v=6Fanea5kcfg | 743,000 |
| Buzzfeed Hates Men | h3h3 | Buzzfeed/Tania Safi | 13:00 | https://www.youtube.com/watch?v=T77frLL_bsg | 12.8 Million |
| Political Correctness—The FALL of MTV | Chris Ray Gun | Francesca Ramsey | 11:33 | https://www.youtube.com/watch?v=E2Jz-tV5Fk | 702,000 |
| EVERYONE IS TRANSPHOBIC | Repzjon | Milo Stewart | 13:27 | https://www.youtube.com/watch?v=4ZpEjw9ahr1 | 363,000 |
| Stefan Molyneux's Native American Genocide—A Response | Shaun | Stefan Molyneux | 36:44 | https://www.youtube.com/watch?v=Xd_nVCWPgiA | 605,000 |
| The Sarkeesian Effect: A Measured Response | hbombberguy | Davis Aurimi and Jordan Owen | 6:54 | https://www.youtube.com/watch?v=nsdlHK8O5yo | 693,000 |

for networked harassment. By utilizing debating techniques, creators portrayed themselves as defenders of enlightened public discourse and their targets as irrational and inaccurate. As we will show, this framing—constructed through selective editing—is not only misleading but also obscures other important dynamics. Specifically, creators interpellate the viewer as part of a networked audience that shares their moral values. This shared ethical baseline enables creators to treat their targets as synecdoches of leftist principles that violate the belief systems of the networked audience, stoking moral outrage, and justifying harassment. By analyzing the networked audiences' use of YouTube's affordances, we find evidence that these videos often lead to networked harassment.

Overall, we identify three distinct forms of amplification linked to networked harassment. First, response videos identify a "target" for the creator's audience and provide a justification for harassing them (amplifying the *potential for* harassment); and second, YouTube affordances facilitate a wide range of opportunities for collaborative harassment against the target (amplifying *avenues for* harassment). Finally, when networked individuals acting on the implicit or explicit justification for harassment presented by a response video make use of these affordances, they collectively amplify the *size and scope* of harassment, even as any individual may only contribute a small amount or not be aware of the broader impact.

Staging a Faux Debate

In our corpus, creators drew on various techniques to simulate the "call and response" structure of an idealized debate. In addition to editing their videos to intercut the target's comments with their own commentary, creators often address their targets in the second person as if they were present. Moreover, creators regularly used a range of argumentative techniques which we argue delegitimizes the (absent) interlocutor. These rhetorical tactics construct the adversary as a soft target against whom harassment would be justified.

First, we found several examples of creators presenting external research to support their arguments and claim superior evidence. For example, Hunter Avallone, in his response video to the kid-friendly channel "Queer Kids Stuff," refers to a 2017 *Independent* article to argue that it is morally wrong to introduce children to concepts like gender transition.² Creators also use external sources to suggest that their targets are misrepresenting or even cheapening scientific inquiry. In a response video targeting nonbinary Muslim drag queen Amrou Al-Kadhi, Paul Joseph Watson claims that his target abuses "physics concepts to give pseudoscientific credibility to a lifestyle choice." Watson's invocation of scientific concepts does not simply paint Al-Kadhi's video as misguided, but as willfully "abusing" science to justify a deviant lifestyle "choice."

Second, creators drew on personal experience to refute their target's claims and offer a frame that is more sympathetic to a creator's ideological position. Blaire White, responding to an argument by another transgender YouTuber, Riley Dennis, that misgendering constitutes a form of violence, offers:

I was working at a clothing store and I had a co-worker who was just intentionally trying to be a bitch, so she would call me a ‘he’ just to, like, ruffle my feathers. But I’ve also experienced getting hit in the face at a concert before. One was violence, one wasn’t . . . I don’t have time to be accused of violence because I forgot to call you “zir.”

Comments such as these do not simply refute an interlocutor’s point, but offer an alternative framework for assessing the target’s statements. White’s frame allows Dennis’s claims to be dismissed as overly sensitive. This creates a “correct” and “incorrect” way of understanding and responding to transmisogyny, one that implicitly places Riley outside the “appropriate” frame.

Third, creators endeavored to tilt the moral landscape in their favor by rejecting targets’ argumentative premises. Beyond rebutting a target’s claims, this rhetorical maneuver opens new axes of argumentation in line with the moral positioning of the creator and their imagined audience. Returning to Watson’s response to Al-Kadhi, he rejects their assertion that there is no necessary contradiction between being both Muslim and queer:

Well I can see why you don’t live in Iraq where gay people are still the victims of honor killings, but you do live in a country where literally 50% of the Muslim population wants to lock you up in a cage for being gay, and yet here you are on the UK’s national broadcaster . . . I think you’ve got your priorities a little bit mixed up there.

On one level, this argument draws attention away from (and downplays) the existence of homophobia in “Western” societies by invoking a common Islamophobic trope. But by suggesting that Al-Kadhi’s “got [their] priorities a little bit mixed up,” it also serves to suggest that the target is an irrational and untrustworthy source.

Response videos position the video creator as a rational debater calmly addressing seemingly misguided, malicious, or emotional appeals. By mimicking debates, they frame their videos as a space of liberal agency, one in which, as Sharon Marcus (2019) describes, rational arguments—“logical, nonviolent, evidence-based debate[s]”—are made with the goal of “producing a consensus with which all members more or less agree” (p. 50). However, we find that framing response videos as instances of rational debate are misleading, as creators frequently selectively edited videos or used dubious external claims to portray targets as irrational and inaccurate. Just as importantly, the creator responds directly to the target’s clips, but the target is not actually present to address such criticisms—thus, the creator always gets the literal and figurative “last word.” We argue that this framing obscures other dynamics at work in these videos, which, as we explore below, stoke moral outrage and promote the shared values of creator and audience.

Constructing a Shared Morality

Response videos presume that viewers are part of a network with shared norms and values. In our sample, creators frequently asked rhetorical questions or addressed the

audience directly. For example, in a video targeting trans vlogger Gigi Gorgeous, alt-right creator James Allsup asks sardonically, “Does this sound like a stable person to you?” Another creator, Andy Warski, responded to a BuzzFeed video about racially insensitive costumes by angrily accusing the feminist vlogger Francesca Ramsey of racism against white people. Talking directly to the audience, he incredulously asks, “Everyone, did you hear that? Did you hear what she just said?” after which he offers his own assessment, calling her a “f*cking dumb b*tch.”

Creators further implicate their audiences by positioning them as targets of progressive critique. While criticizing trans vlogger Milo Stewart, vlogger Repzion tells the viewer “Today, you’re gonna be called misogynist, you’re gonna be called racist and you’re gonna be called transphobic.” Here, Repzion works off the premise that anyone watching Stewart’s video would, like him, object to the general claim that all cisgender people are transphobic. Rather than interpreting it as a structural critique, Repzion directs his followers to see it as a personal attack. This form of direct address interpellates the audience as a sympathetic participant in the critique of the target, as well as someone whose beliefs and experiences are similar to the creator’s.

Moreover, the use of offensive language throughout these videos presumes that the viewer is not “politically correct.” The ableist slur “ret*rded” was found frequently in the sample, along with gendered slurs and insults, including “c*nt,” “p*ssy,” “dumb b*tch,” and “old hag,” and, less often, transphobic and homophobic slurs. The prevalence of such language in the corpus suggests that vloggers presume their audience does not find it offensive, and in fact may consider the use of these words a rebellion against so-called leftist censorship.

By constructing a networked audience and painting their “interlocutor” as irrational, vloggers position their target as violating norms shared between creator and audience. These norm violations can serve as justifications for networked harassment, including the idea that leftists are the “real” racists because they are racist against white people, that feminism is an antimale belief system, that LGBTQ people harm children by promoting gender confusion, and that social justice is counterproductive by hurting the very people it claims to help, all of which appeared in our corpus. In these cases, the vlogger often stands in for “the left” as a whole.

We can see these dynamics at work in “It’s OK to celebrate Thanksgiving,” in which the vlogger No Bullshit targets Imaeyen Ibanga, a Nigerian-American video producer at Al Jazeera’s video network AJ+. The original video, “Thanksgiving, What You Need to Know,” interviews Native American activists to argue that dominant narratives of Thanksgiving are historically inaccurate. No Bullshit’s video erroneously characterizes the AJ+ video as an argument for ending Thanksgiving celebrations. When Ibanga says “This Sunday we’re exploring the real story of Thanksgiving,” No Bullshit responds, “I think you mean you’re destroying Thanksgiving, at least you’re trying to,” portraying her argument as an obviously unpopular position. Throughout his critique, No Bullshit labels AJ+ and Ibanga “un-American” and “anti-White.” The audience is thus encouraged to defend themselves, Thanksgiving, and America itself from attacks coming from Black and Native women.

This blueprint for harassment is echoed in the majority of videos in our corpus. By mixing ad hominem attacks with misrepresentation and criticism of left-wing beliefs, creators stoke moral outrage in their networked audience by suggesting the vlogger is attacking or harming White people, men, children, or America, and provide a list of talking points that can be repeated when harassing a target. Thus, harassment is implicitly justified as righting a moral wrong, reinforcing the norms of the networked audience. Taken together, we argue that these phenomena—the staging of a faux “rational” debate while simultaneously invoking moral outrage among like-minded viewers—offers conditions for networked harassment even when it is not called for or is even repudiated by creators.

In this way, the creators whose content we analyzed acted as amplifiers for networked harassment. That is, they exposed their audiences to a target and suggested the target violated moral norms shared by those audiences. Importantly, even if creators explicitly ask their audiences not to engage in harassment, our findings suggest that the response videos in our corpus still created conditions that seeded the potential for networked harassment by providing the blueprint discussed above.

Affording Harassment

While creators thus amplify blueprints for harassment, we found that YouTube’s social affordances facilitate the actual avenues of networked harassment. For example, each video uploaded to YouTube has an “About” section positioned directly below the video; of the 15 videos we sampled, seven included a link to the target video in this section. While their inclusion could be justified as a kind of citational practice, encouraging viewers to hear the target’s arguments for themselves, they also create a direct link to the target’s channel. We suggest that this connection between videos functions as a vector for networked harassment, as the affordances of YouTube funnel a creator’s audience toward their target’s profile. Our point is less that these links cause harassment than that they may remove a barrier to doing so.³

On arriving at a targeted video, users are afforded several mechanisms to interact with the target. Most frequently, we found that users leverage engagement features to express their displeasure with the targeted creator. By default, YouTube includes a “like” and “dislike” metric for each video. These metrics often serve as a marker of status among YouTubers, but also provide immediate feedback for creators about audience preference. Four of the videos we examined, all of which were directly linked in a response video’s “About” section, had an extremely high ratio of dislikes to likes (Table 2). This suggests these videos were subject to “vote brigading,” a long-standing practice among internet subcultures of collectively skewing the results of online voting and review systems (e.g., Dosono & Semaan, 2020; Raftery, 2016). (YouTube itself is aware of this practice on its own platform, euphemistically describing the phenomenon as “dislike mobs”; Alexander, 2019b). All the response videos, by contrast, had a high ratio of likes to dislikes, suggesting that the audience for response videos is primarily like-minded viewers. These discrepancies have important implications for creators on the platform: the metric is treated as a status marker, and creators

Table 2. Likes to Dislikes Ratio on Response Video and Targeted Video.

| Response video | Ratio of likes/dislikes | Targeted video likes/dislikes |
|--|-------------------------|--|
| Sensitive Tranny Thinks Words Are Violent | 31,000/683 | 1,600/15,000 |
| Dear Rebecca Watson | 6,300/263 | 3,600/6,400 |
| Feminism vs. FACTS | 29,000/2,700 | 8,200/14,000 |
| It's Okay To Celebrate Thanksgiving | 6,500/260 | 409/1,600 |
| FEMINISM IS FOR POOPYHEADS | 67,000/1,500 | No link |
| Gay Muslim Drag Queen Causes Confusion | 55,000/679 | No link |
| Buzzfeed Hates Men | 506,000/7,900 | Comments and likes/dislikes turned off on targeted video |
| Everyone is Transphobic | 14,000/1,200 | Targeted video deleted |
| The FALL of MTV | 29,000/441 | No link |
| Stefan Molyneux's Native American Genocide | 19,000/1,000 | 14,000/5,000 |
| The Sarkeesian Effect | 17,000/1,000 | No link |

have long theorized that a high dislike to like ratio may prevent YouTube from recommending a video in its algorithms.

We found that users also participated in harassment in the “Comments” section of targeted videos. Users’ remarks ranged from the generically rude (e.g., “So much cringe,” on the video targeted by No Bullshit) to barely veiled threats of physical harm (e.g., “Oh Riley. You would be having such a great time here, in Poland. Lots of opportunities to find out what violence means,” on the video targeted by Blaire White). Many comments echo the language and arguments of creators, suggesting that creators may provide a kind of script for subsequent harassment. For example, one user writes “What do you C*NTS know about other cultures?!” on the video targeted by Andy Warski, echoing one of Warski’s statements and most-used slurs. Regardless of when the original video was uploaded, we found that negative comments cluster around the time that the response video targeting it was uploaded. Furthermore, in each of the targeted videos in our sample, the comments with the most likes were clearly antagonistic in nature, drowning out any countercommentary from the targeted YouTuber’s own audience. Additionally, we found very few instances of critical commentary on the “response videos” from users who do not share the responding YouTuber’s ideological valence, reflecting the asymmetry of this dynamic. While there were similar numbers of comments on the reply videos as on the targeted videos, they consistently displayed strong qualitative differences. A full content analysis of the comment sections on all videos is outside the scope of this article, but we analyzed a brief case study which suggests a stark discrepancy. Focusing on Blaire White’s response video and the targeted video by Riley Dennis, we performed a preliminary content analysis of comments and found

Table 3. Comment Metrics on Sample Response Video and its Target.

| Video | Number of exportable comments | Critical comments | Supportive comments |
|---|-------------------------------|-------------------|---------------------|
| Response video from corpus: Sensitive Tranny Thinks Words Are Violent | 3,443 | 110 (3%) | 2,779 (81%) |
| Targeted video: Misgendering trans people is an act of violence | 1,390 | 1,185 (85%) | 75 (5%) |

that, while only 3% of the comments on Blaire White’s video were negative or adversarial in nature, 85% of those on Dennis’s video were (see Table 3). Furthermore, several of the comments on Dennis’s video directly referenced which response videos had inspired them to come to this page (including White’s video).

Taken together, our findings indicate not only that the response video format provides a blueprint for networked harassment, but that networked audiences use the social affordances of YouTube to act on that blueprint, whether the creator explicitly promotes this or not. For this reason, we refer to them as *harassment affordances*, even as those same affordances may be used for more general purposes. As Davis and Chouinard (2016) suggest, technological artifacts can request, demand, allow, encourage, discourse, and refuse particular actions. Moreover, technical features *underdetermine* design outcomes, meaning that these mechanisms of affordance are not enacted merely by an artifact’s particular features, but also by their users and their contexts for practice. In the context of response videos, we are not suggesting that the platform’s features *demand* harassment, but, rather, that these affordances can allow or even encourage harassment under certain circumstances—namely, when amplifiers and audiences wish to engage. In other words, our argument presumes a combination of techno-social forces working in tandem to produce conditions amenable to harassment. Our use of “harassment affordances” is meant to highlight this ambivalence by refusing to reduce harassment to a matter of YouTube’s technical features even as we highlight them, because the same features may afford different design outcomes depending on the users and their context.

Our findings indicate that identifying and observing such dynamics requires a holistic view of networked audiences and creators that goes beyond analysis of individual videos and points to the dynamics around videos and the platform architectures that support them. While response video creators amplified the potential for harassment of a given target, the affordances of social media facilitated a more diffuse amplification of harassment distributed across a broad group of people. Viewers collectively made use of those affordances, shaping them to their needs. The result of this socio-technical assemblage is an amplification of the size, scope, and methods of harassment toward the target across a range of avenues.

Notably, since we conducted our original analysis, four videos in our corpus were removed by YouTube. While this may initially seem to indicate progress on YouTube’s

treatment of harassment, we argue it still reflects a lack of consistency and transparency. It is unclear whether the videos were removed for instigating harassment, for hate speech, or for another reason. Additionally, of the four removed videos in our corpus, two of them were removed when the creator's entire channel was suspended, while the other two were only penalized as individual videos. Finally, it is unclear whether any of the other videos in our corpus were demonetized, because that information is not made public by the platform (Caplan & Gillespie 2020). Thus, the actions taken by YouTube seem to simultaneously reflect a knowledge of abusive behavior while simultaneously demonstrating inconsistency and opacity in their responses.

Leftist Response Videos

While our analysis primarily focused on right-wing videos, we analyzed two left-wing videos for exploratory comparison. While this sample is too small to make meaningful claims about broader patterns of left-wing YouTube videos, these two videos differed from each other and from anti-SJW creators in their approach. The first video comes from a YouTuber known as Shaun, whose channel largely consists of response videos addressing right-wingers. Shaun painstakingly responds to White nationalist Stefan Molyneux's "The Truth About the Native American Genocide" video. He cites archival and scholarly sources, explaining historical events like the Sand Creek Massacre and the Treaty of Fort Wise. In this sense, Shaun's video stayed more faithful to the faux-staged debate format than the right-wing videos in our corpus; at no point, however, does he appeal to the emotions of his viewers or a shared sense of morality between them.

This is also reflected in the comments. While Shaun linked to Molyneux's video and his viewers left comments, they largely adhere to the debate style employed by Shaun. One says, "I recommend Shaun's response titled "Stefan molyneux's native American genocide—a response" anyone who is actually interested in other points of view and not just satisfying their ego. Personally I agree with Shaun on the topic but I'll leave the judgments up to you." Another says "Your research is one sided and leaves out the Native side of the history. Nothing mentioned of Sand creek massacre or Wounded Knee massacre." Other comments are similarly solemn and factual. Additionally, there appears to have been little or no manipulation of the likes/dislikes feature, as the original video's ratio of likes to dislikes remains quite high.

The second video we analyzed was produced by the left-wing video essayist Hbomberguy. His response to Davis Aurini and Jordan Owen more closely fits the pattern we observed in right-wing videos but differs in important ways. He lobs personal attacks against Owen and Aurini, calling them "fat, pasty White guys," and insulting their sexual prowess. However, Hbomberguy does not engage with their talking points directly, position his targets as dangerous, or as synecdoches of larger right-wing thought. Rather, he treats their claims as self-evidently ridiculous. Furthermore, he does not link to Aurini's or Owen's videos, thus not making use of any affordance that might encourage his viewers to partake in harassment.

Thus, our preliminary comparison suggests such videos do not create the "blueprint for harassment" we observed in right-wing videos. This limited comparison is

not meant to make a generalizable claim that harassment does not exist within left-wing online communities, nor do we suggest that online cultures are the determinant factor in networked harassment. Instead, this comparison helps illuminate our broader techno-social argument that a symbiotic relationship often exists between amplifiers, viewers, and platform architecture that leads to networked harassment, and that this symbiosis exists among anti-SJW response video creators and their viewers on YouTube.

Discussion: Platform Accountability

Our findings enable us to revisit YouTube's response to Carlos Maza with greater understanding. Responding to Maza's claims, YouTube's Chris Dale (2019) blogged, "we consider the context of the entire video . . . we take into consideration whether criticism is focused primarily on debating the opinions expressed or is solely malicious." A video that criticizes another creator in the context of a larger political debate—that is, a response video—does not, therefore, violate YouTube's policy (Goldman, 2019). Maza was incensed by this reply. He told Gimlet's Reply All podcast that

what YouTube has done is basically say, "Hey if you want to get away with hate speech on the platform just make sure you sandwich it between, you know, ten minutes of political nonsense that isn't that [harassment]." It's not an anti-harassment policy, it's an instruction manual for how to harass. (Goldman, 2019).

In line with Maza's critique, our findings indicate that YouTube's harassment policies are insufficient because they (a) take notions of "debate" and "opinions" at face value, ignoring the strategic work these frames do in setting *blueprints for harassment* while obfuscating blame; (b) treat videos in isolation, rather than accounting for the broader ecosystem of networked audiences and amplifiers, and the shared moral values these networks create and defend; and (c) fail to account for the platform's *harassment affordances* that offer opportunities for harassment and can be weaponized by creators and their audiences against targets.

While our findings indicate that amplifiers play a major role in inciting harassment on the platform, it is often difficult to conclusively link a creator with subsequent harassment of their target, given that multiple right-wing YouTubers often respond to the same video at similar times and audience members rarely mention the motivations behind their attacks. This poses a challenge from an accountability perspective given the plausible deniability that creators are granted for actions undertaken by their fans. Current platform policies are unable to grapple with the fundamentally networked nature of this dynamic, in which amplifiers and audiences work together, but neither is solely responsible for, or even aware of, the impact on targets.

Finally, our findings indicate not only that platform *policies* are insufficient, but that *enforcement* of these policies is problematic. Specifically, there is evidence that platforms are less likely to discipline an "influencer" than a less popular account. *The*

Washington Post found that YouTube content moderators claim the platform doles out “more lenient punishments for top video creators whose work brings in the most money for the company” (Dwoskin, 2019). In this sense, there are inconsistencies with how YouTube’s terms of service are enforced and there is no transparency around how or when they remove creators from their platform. While YouTube currently gives more leeway to its biggest and most profitable creators, our findings indicate that, as amplifiers, popular creators have the possibility of stoking enormous amounts of harassment.

Conclusion

YouTube has long positioned itself as an egalitarian space to entice creators into joining this digital “public square” (Gillespie, 2010). This is misleading because broadcasting and response videos differ greatly from idealistic visions of debate and opinion sharing. The assumption that the best ideas rise to the top of public discourse ignores the long history of media’s use for fascist and authoritarian propaganda and demagoguery, which generally appeal to viewers’ emotions and sense of identity, rather than empirical evidence and logical reasoning. Indeed, despite presenting themselves as cool-headed rationalists, we found that the creators we analyzed often appealed to emotions and personal attacks, throwing insults and slurs while displaying (sometimes extreme) anger alongside generic right-wing bromides.

Our findings add to the growing body of scholarship moving beyond dyadic models of harassment that presume one abuser and one victim. Instead, response videos provide a valuable example of harassment taking place through distributed networks, generally inspired by high-profile creators who amplify a potential target and provide justification for their harassment, purposefully or not. Our findings also add to research that takes a critical approach to participatory culture, analyzing its potential harmful effects and complicating original conceptions of its inherently democratizing or progressive nature. Finally, we build on literature that argues that platform affordances play a role in shaping destructive cultures online. Much like Adrienne Massanari (2015) argued that Reddit’s architecture and voting systems encouraged the growth of “toxic technocultures” on the platform, we argue that YouTube’s social affordances frequently serve as harassment affordances.

Our analysis points to several promising areas for potential future research. While this study focused on anti-SJW response videos, we suspect similar dynamics may exist within other communities and platforms. How might amplification interact with networked harassment on other platforms with different political communities and features? Do harassment affordances exist on other platforms? How do platform policies account for, or fail to account for, such dynamics?

Overall, this article demonstrates that YouTube’s standard for harassment is unrealistically narrow. Response videos *amplify* harassment by highlighting a specific target, creating a moral justification for harassment, and providing a blueprint for how members of the networked audience should interact with the target. Without addressing the

role of amplification, networked audiences, and harassment affordances, YouTube's policies will remain insufficient for grappling with current harms on the platform.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by a Stanford Graduate Fellowship, a grant from Luminare, and a Faculty Fellowship at the Institute of Arts and Humanities at the University of North Carolina at Chapel Hill.

Notes

1. The term affordance has a complex history, but generally describes how technical objects enable and constrain users on a spectrum that runs from demanding particular actions to outright refusing them (Davis & Chouinard, 2016). Conceptually, "affordance" cleaves between constructivist and determinist arguments about the effects of technological artifacts, recognizing that nonhuman artifacts have agency but are still shaped by context.
2. Such citations are often misleading. While the *Independent* article Avallone cites notes that some patients seek gender reversal surgery, it also states "there are no statistics regarding reversal surgeries in the UK."
3. Our corpus does not address whether specific users who engaged in harassing behaviors did so through a provided link to a target. Moreover, audiences may engage in harassment against the target on multiple platforms beyond YouTube.

References

- Abidin, C. (2015). Communicative ♥ Intimacies: Influencers and Perceived Interconnectedness. *Ada: A Journal of Gender, New Media & Technology*, 2015(8). <http://adanewmedia.org/2015/11/issue8-abidin/>
- Abidin, C. (2018). *Internet celebrity: Understanding fame online*. Emerald Publishing Limited.
- Alexander, J. (2019a, May 31). YouTube investigating right-wing pundit Steven Crowder for harassing Vox.com host. *The Verge*. <https://www.theverge.com/2019/5/31/18647621/youtube-steven-crowder-bullying-harassment-twitter-vox-carlos-maza>
- Alexander, J. (2019b, February 1). YouTube wants "dislike mobs" to stop weaponizing the dislike button: One option is to do away with it altogether. *The Verge*. <https://www.theverge.com/2019/2/1/18207189/youtube-dislike-attack-mob-review-bomb-creator-insider>
- Amin, S. (2019, July 2). *Can the left win YouTube?* *The New Republic*. <https://newrepublic.com/article/154399/can-left-win-youtube>
- Banet-Weiser, S., & Miltner, K. M. (2016). # MasculinitySoFragile: Culture, structure, and networked misogyny. *Feminist Media Studies*, 16(1), 171-174. <https://doi.org/10.1080/14680777.2016.1120490>
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation and radicalization in American politics*. Oxford University Press.

- Berryman, R., & Kavka, M. (2017). "I Guess A Lot of People See Me as a Big Sister or a Friend": The role of intimacy in the celebrification of beauty vloggers. *Journal of Gender Studies*, 26(3), 307-320. <https://doi.org/10.1080/09589236.2017.1288611>
- Blackwell, L., Chen, T., Schoenebeck, S., & Lampe, C. (2018). When online harassment is perceived as justified. In *Proceedings of the 12th International Conference on Web and Social Media* (pp. 22-31). AAAI. https://yardi.people.si.umich.edu/pubs/Schoenebeck_RetributiveJustice18.pdf
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(28), 7313-7318. <https://doi.org/10.1073/pnas.1618923114>
- Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*. John Wiley.
- Burgess, J., & Matamoros-Fernández, A. (2016). Mapping sociocultural controversies across digital media platforms: One week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, 2(1), 79-96. <https://doi.org/10.1080/22041451.2016.1155338>
- Caplan, R., & Gillespie, T. (2020). Tiered governance and demonetization: The shifting terms of labor and compensation in the platform economy: *Social Media + Society*. Advance online publication. <https://doi.org/10.1177/2056305120936636>
- Chess, S., & Shaw, A. (2015). A conspiracy of fishes, or, how we learned to stop worrying about# gamergate and embrace hegemonic masculinity. *Journal of Broadcasting & Electronic Media*, 59(1), 208-220. <https://doi.org/10.1080/08838151.2014.999917>
- Citron, D. (2014). *Hate crimes in cyberspace*. Harvard University Press.
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11), 769-771. <https://doi.org/10.1038/s41562-017-0213-3>
- Cunningham, S., & Craig, D. (2017). Being "really real" on YouTube: Authenticity, community and brand culture in social media entertainment. *Media International Australia*, 164(1), 71-81. <https://doi.org/10.1177/1329878X17709098>
- Dale, C. (2019, June 5). *Taking a harder look at harassment*. Official YouTube Blog. <https://youtube.googleblog.com/2019/06/taking-harder-look-at-harassment.html>
- Davis, J. L., & Chouinard, J. B. (2016). Theorizing affordances: From request to refuse. *Bulletin of Science, Technology & Society*, 36(4), 241-248. <https://doi.org/10.1177/0270467617714944>
- Dosono, B., & Semaan, B. (2020). Decolonizing tactics as collective resilience: Identity work of AAPI communities on Reddit. In *Proceedings of the ACM on Human-Computer Interaction* (1-20). <https://doi.org/10.1145/3392881>
- Duffy, B. E. (2017). *(Not) getting paid to do what you love: Gender, social media, and aspirational work*. Yale University Press.
- Dwoskin, E. (2019, August 9). YouTube's arbitrary standards: Stars keep making money even after breaking the rules. *The Washington Post*. <https://www.washingtonpost.com/technology/2019/08/09/youtubes-arbitrary-standards-stars-keep-making-money-even-after-breaking-rules/>
- Fagerjord, A. (2010). After convergence: YouTube and remix culture. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 187-200). Springer.
- Finkelhor, D., Mitchell, K. J., & Wolak, J. (2000). *Online victimization: A report on the nation's youth* (No. 6-00-020). National Center for Missing & Exploited Children.
- Gillespie, T. (2010). The politics of "platforms." *New Media & Society*, 12(3), 347-364. <https://doi.org/10.1177/1461444809342738>

- Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.
- Goldman, A (host). (2019, July 11). Louder. [Audio podcast episode]. In *Reply all*. Gimlet Media. <https://gimletmedia.com/shows/reply-all/rnhzlo>
- Henry, N., & Powell, A. (2015). Beyond the “sext”: Technology-facilitated sexual violence and harassment against adult women. *Australian & New Zealand Journal of Criminology*, 48(1), 104-118. <https://doi.org/10.1177/0004865814524218>
- Hill, C., & Johnson, H. (2020). Online interpersonal victimization as a mechanism of social control of women: An empirical examination. *Violence Against Women*, 26(12-13), 1681-1700. <https://doi.org/10.1177/1077801219870608>
- Hine, G., Onalapo, J., De Cristofaro, E., Kourtellis, N., Leontiadis, I., Samaras, R., Stringhini, G., & Blackburn, J. (2017). Kek, cucks, and God Emperor Trump: A measurement study of 4chan’s Politically Incorrect forum and its effects on the web. In The Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM-17): 92-101. AAAI.
- Jane, E. A. (2014). “Back to the kitchen, cunt”: Speaking the unspeakable about online misogyny. *Continuum*, 28(4), 558-570. <https://doi.org/10.1080/10304312.2014.924479>
- Jhaver, S., Chan, L., & Bruckman, A. (2018). The view from the other side: The border between controversial speech and harassment on Kotaku in action. *First Monday*, 23(2). <http://firstmonday.org/ojs/index.php/fm/article/view/8232/6644>
- Klonick, K. (2017). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598-1669. <https://harvardlawreview.org/2018/04/the-new-governors-the-people-rules-and-processes-governing-online-speech/>
- Lange, P. G. (2007, March 31). *Commenting on comments: Investigating responses to antagonism on YouTube* [Paper presentation]. Annual Conference of the Society for Applied Anthropology, Tampa, FL. https://www.researchgate.net/publication/228615792_Commenting_on_Comments_Investigating_Responses_to_Antagonism_on_YouTube
- Lange, P. G. (2009). Videos of affinity on YouTube. In P. Snickars & P. Vonderau (Eds.), *The YouTube reader* (pp. 228-247). National Library of Sweden.
- Lenhart, A., Ybarra, M. L., Zickuhr, K., & Price-Feeney, M. (2016). *Online harassment, digital abuse, and cyberstalking in America*. Data & Society Research Institute. https://www.data-society.net/pubs/oh/Online_Harassment_2016.pdf
- Lewis, R. (2018). *Alternative influence: Broadcasting the reactionary right on YouTube*. Data & Society Research Institute. https://datasociety.net/wp-content/uploads/2018/09/DS_Alternative_Influence.pdf
- Madden, S., Janoske, M., Winkler, R. B., & Edgar, A. N. (2018). Mediated misogynoir: Intersecting race and gender in online harassment. In J. R. Vickery & T. Everbach (Eds.), *Mediating misogyny: Gender, technology, and harassment* (pp. 71-90). Palgrave Macmillan.
- Marcus, S. (2019). *The drama of celebrity*. Princeton University Press.
- Martinez, N. (2019, April 9). *Study: Facebook is still not censoring conservatives*. Media Matters for America. <https://www.mediamatters.org/facebook/study-facebook-still-not-censoring-conservatives>
- Marwick, A. (2015). You may know me from YouTube: (Micro-)Celebrity in Social Media. In *A companion to celebrity* (pp. 333-350). John Wiley.
- Marwick, A., & boyd, d. (2014). “It’s just drama”: Teen perspectives on conflict and aggression in a networked era. *Journal of Youth Studies*, 17(9), 1187-1204. <https://doi.org/10.1080/13676261.2014.901493>
- Marwick, A., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies*, 18(4), 543-559. <https://doi.org/10.1080/14680777.2018.1450568>

- Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society Research Institute. https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf
- Massanari, A. (2015). # Gamergate and the fapping: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3), 329-346. <https://doi.org/10.1177/1461444815608807>
- Moloney, M. E., & Love, T. P. (2018). Assessing online misogyny: Perspectives from sociology and feminist media studies. *Sociology Compass*, 12(5), e12577. <https://doi.org/10.1111/soc4.12577>
- Newton, C. (2019, April 11). *The real bias on social networks isn't against conservatives*. *The Verge*. <https://www.theverge.com/interface/2019/4/11/18305407/social-network-conservative-bias-twitter-facebook-ted-cruz>
- Otoni, R., Cunha, E., Magno, G., Bernardina, P., Meira, W., & Almeida, V. (2018). Analyzing right-wing YouTube channels: Hate, violence and discrimination. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 323-332). ACM.
- Papacharissi, Z. (2016). Affective publics and structures of storytelling: Sentiment, events and mediality. *Information, Communication & Society*, 19(3), 307-324. <https://doi.org/10.1080/1369118X.2015.1109697>
- Pater, J. A., Kim, M. K., Mynatt, E. D., & Fiesler, C. (2016). Characterizations of online harassment: Comparing policies across social media platforms. In *Proceedings of the 19th International Conference on Supporting Group Work* (pp. 369-374). ACM.
- Pihlaja, S. (2014). *Antagonism on YouTube: Metaphor in online discourse*. Bloomsbury.
- Raftery, B. (2016, September 14). IMDb voters are tanking Indies before they're even released. *Wired*. <https://www.wired.com/2016/09/imdb-voters-tanking-indies-theyre-even-released/>
- Raun, T. (2018). Capitalizing intimacy: New subcultural forms of micro-celebrity strategies and affective labour on YouTube. *Convergence*, 24(1), 99-113. <https://doi.org/10.1177/1354856517736983>
- Ribeiro, M. H., Otoni, R., West, R., Almeida, V. A. F., & Meira, W. (2020). Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 131-141). ACM.
- Sobieraj, S. (2018). Bitch, slut, skank, cunt: Patterned resistance to women's visibility in digital publics. *Information, Communication & Society*, 21(11), 1700-1714. <https://doi.org/10.1080/1369118X.2017.1348535>
- Soldner, F., Ho, J. C., Makhortykh, M., Van der Vegt, I., Mozes, M., & Kleinberg, B., (2019). Uphill from here: Sentiment patterns in videos from left- and right-wing YouTube news channels. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science* (pp. 84-93). Association for Computational Linguistics.
- Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*.
- TeamYouTube. (2019, June 4). *As an open platform, it's crucial for us to allow everyone—from creators to journalists to late-night TV hosts—to express their opinions w/in the scope of our policies. Opinions can be deeply offensive, but if they don't violate our policies, they'll remain on our site.* [Tweet]. Twitter. <https://twitter.com/TeamYouTube/status/1136055805545857024>
- The YouTube Team. (2013, August 27). *So long, video responses . . . Next up: Better ways to connect*. YouTube Creator Blog. <https://youtube-creators.googleblog.com/2013/08/so-long-video-responsesnext-up-better.html>

- Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior, 26*(3), 277-287.
- Vitak, J., Chadha, K., Steiner, L., & Ashktorab, Z. (2017). Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 1231-1245). ACM. <https://dl.acm.org/doi/10.1145/2998181.2998337>

Author Biographies

Rebecca Lewis is a stanford graduate fellow and a PhD candidate in the Department of Communication at Stanford University. Her research focuses on digital political subcultures, internet celebrities, and online video platforms.

Alice E. Marwick is an associate professor in the Department of Communication and a principal researcher at the Center for Information, Technology and Public Life at the University of North Carolina at Chapel Hill.

William Clyde Partin is a PhD candidate in the Department of Communication at the University of North Carolina at Chapel Hill and a researcher at the Data & Society Research Institute.